

# Discursive analysis of itineraries in a historical regional corpus of travels: syntax, semantics, and pragmatics in a unified type theoretical framework

**Richard Moot\***

Université de Bordeaux

LaBRI(CNRS)

SIGNES(INRIA)

moot@labri.fr

**Laurent Prévot**

Université de Provence

Laboratoire Parole et Langage

laurent.prevot@lpl-aix.fr

**Christian Retoré\***

Université de Bordeaux

LaBRI(CNRS)

SIGNES(INRIA)

retore@labri.fr

## 1 Introduction

In this paper we will discuss the application of (Segmented) Discourse Representation Theory and the Generative Lexicon to the analysis of a historical French corpus of itineraries in the Pyrénées. Our research will focus in particular on how type coercion (Pustejovsky, 1995) can help us give a correct analysis of cases of so-called “fictive motion” (Talmy, 1999), which is evident in phrases like.

- (1) The road runs along the coast for two hours.
- (2) The path descended abruptly.

This case is particular in that an entity (which is considered immobile and which, in the context, defines a *path*) is the subject of a movement verb and that the combination is interpreted as a generic statement about the nature of this path, without any movement necessarily taking place.

## 2 Context

The context of the current research project is to provide a semantic representation the paths traversed or described by the authors of the different books in a corpus of itineraries in the Pyrénées mountains. It is a 19th century corpus consisting of 576.334 words, containing (among many other details and descriptions) narratives of the routes followed and the places visited by the authors. Other information about the corpus and its spatio-temporal analysis can be found in (Loustau, 2008; Asher et al., 2008).

The problem on which we will focus here is that in many cases, non-agentive movement verbs, such as “descendre” (corresponding to *go/come*

*down* in English) can take as its subject not only a person (sentence (3) below), but also an immobile subject, provided it can be associated (or *coerced*) to a path, such as a fence or a road (sentence (4)).

- (3) Jean descend.
- (4) Le chemin descend.

The phenomenon evident in Examples (1), (2) and (4) has been called “fictive motion”, (Talmy, 1999): our mind’s eye or an imaginary traveler moves along a stationary object such as a road, a fence or a table. It is our goal to give an account of this phenomenon in a type-theoretic framework.

The examples below, which are taken from our corpus, show some further particularities of fictive motion.

- (5) Nous coupons ici un sentier qui vient du port de Barroude (...)  
*Here, we cross a path which comes from the pass of Barroude*
- (6) Plus loin, de nobles hêtres montent sur le versant (...)  
*Further away, noble beeches climb the slope*
- (7) (...) cette route qui monte sans cesse pendant deux lieues  
*this road which climbs incessantly for two miles*

In example (5), it is clear — by the meaning of “couper” (cross) — that the authors do *not* take the path described. In example (6), there is no real path specified and we interpret the sentence as the author’s gaze following a path along the beeches up the slope, whereas in example (7) there is adverbial modification: “incessantly” and “for two miles”. Though the class of permitted adverbs is semantically restricted<sup>1</sup>,

<sup>1</sup>In English we can use manner of motion verbs such as “crawl” and “run” as well

\*Richard Moot and Christian Retoré have benefited from a grant from the Conseil Régional d’Aquitaine, in the context of the Itipy projet and like to thank the other participants of the Itipy project for their discussion on many of the themes of this paper

it is possible to have temporal adverbs such as “the road went along the coast for two hours” and some manner adverbs such as “the path descended abruptly/slowly” which do not commit us to inferring that the author actually took the road. All of this suggests that we can interpret a static object as the (abstract) *process* of traveling along it.

### 3 Syntax and Semantics

Our semantic framework is integrated in a wide-coverage categorial grammar for French (Moot, 2010a), which has been semi-automatically extracted from the French Treebank (Abeillé et al., 2003). The wide-coverage grammar and the Grail parser (Moot, 2010b) combine to parse unseen sentences with a precision comparable to the best-known categorial parsers for English.

Categorial grammars are an especially appropriate choice in the current context because of their tight integration of syntax and semantics: each derivation in a categorial grammar corresponds to a typed-lambda term and this makes writing Montague-style semantics for categorial grammars particularly simple.

As is well-known, however, the possibilities of lambda-calculus semantics permit an integration with more modern theories of semantics, such as discourse representation theory (Kamp and Reyle, 1993) and the Generative Lexicon (Pustejovsky, 1995)

(Bos et al., 2004) and (Moot, 2010b) show that wide-coverage compositional semantics using DRT is possible (for English and for French respectively). In addition (Bassac et al., 2010) show that ideas from the Generative Lexicon can be implemented using polymorphic lambda term assignments to lexical entries.

The current paper proposes an extension to the work of (Moot, 2010b) which permits the system to handle cases of fictive movement.

### 4 Lexical Semantics

Our semantic approach is generated in the tradition of lexical semantics called the Generative Lexicon (Pustejovsky, 1995), especially in its type-logical interpretation of (Bassac et al., 2010)

In our type hierarchy, have two specific types of spatial arguments *regions* and *paths*. Two functions *source* and *destination* convert a path  $p$  to its source region and its destination region<sup>2</sup>. We

<sup>2</sup>We are aware that there are many ways to refer to places

also assume a spatial variable *here* which denotes the position and orientation of the spatial reference point (which does not necessarily correspond to the position of the narrator; in this sense it is closer to a spatial equivalent of the Reichenbachian “reference time” than it is to the constant “now”: in a DRS it is most naturally implemented as a succession of values as is the reference time  $t$ ). Both the position and orientation are necessary to understand a discourse like the following.

- (8) a. My new apartment is awesome.
- b. The entrance hall is spacious.
- c. To the left, there is the living room.

In this discourse, we can make sense of the expression “to the left” only because we make a kind of “virtual visit” with up/down, forward/backward and left/right well-defined.

The distinction between regions and paths is rather standard (Jackendoff, 1983). It is motivated by selectional restrictions on verbs: some verbs, such as “stay + PP” are only grammatical when the PP is a preposition denoting a region argument, whereas other verbs, such as “pass + PP” can only occur with a number of PPs, all of which denote a path. This distinction is muddled slightly by the possibility to coerce a region  $r$  into a path. As is well known, some prepositions, such as “vers” (towards), can — at least in their spatial uses — only denote paths.

For our semantic analysis, we interpret all motion verbs as being relations between one or more entities and a *path*. This argument can be left implicit (ie. when we say “John ran”). Verbs specify lexically which of their arguments follow this path (subject, object or both).

“Le chemin” with type assignment  $np - \iota x^{immobile\_object}.chemin(x)$  does not combine with “descend” which requires a person as its argument, as indicated by its lexical entry which is of the general form  $np \setminus s - \lambda y^{person} \dots$  (to focus the discussion on the coercion mechanism, we give only a schematic entry at this point, the complete entry is shown in Figure 1).

Figure 1 shows the  $\lambda$ -DRTs for both the “le chemin” in its coerced form and “descend”.

Some comments about the lexical semantics. Motion verbs are analysed by the “neutral” predicate *travel* (in the style of (Miller and Johnson-

in the *middle* of the paths as well. However, we assume that this is done by relations rather than by functions.

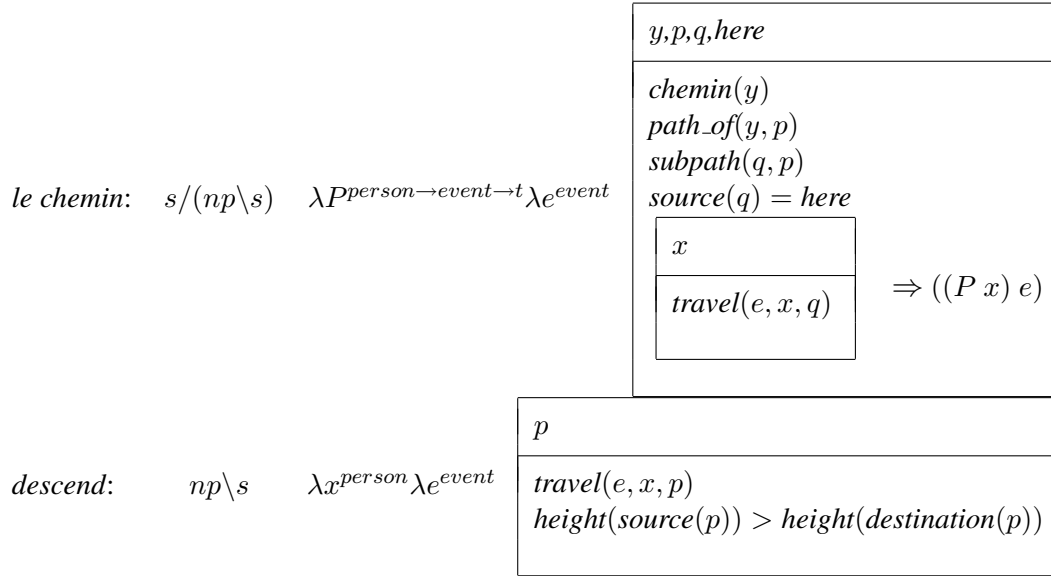


Figure 1: Coerced lambda term for “le chemin” and lambda term for “descend”

Laird, 1976)) which takes an event  $e$ , a moving entity  $x$  and a path  $p$  as its arguments. It is true if the traveler  $x$  follows path  $p$  during event  $e$ . The functions *source* and *destination* are functions from paths to their source and destination regions, whereas *height* is a function from regions to their vertical coordinate. Taken together, the semantic entry for “descend” states therefore that given an argument which is a person  $x$  and an argument which is an event  $e$ , the DRS will verify that there is a path  $p$  such that  $x$  follows  $p$  and that the height at the start of this path  $p$  is greater than his height at the end of it.

The coerced entry for “le chemin” states that given a  $vp$  semantics  $P$  and an event  $e$ , it will assert that there is an entity  $y$  (which is a path in the sense that it extends over a certain amount of space), the predicate *path\_of* will relate  $y$  to one or more directed paths  $p$  and a sub-path  $q$  of  $p$  going forward from *here* is selected. Note that having both  $y$  and  $p$  as referents in the universe of the DRT is necessary to account for modifiers of both aspects of the path, as in “a brick road to Pau”.

This analysis, with the variable *here* having both a place and an orientation, has the pleasant consequence of there being no incoherence between saying “le chemin monte” and “le chemin descend” at exactly the same place but with just the orientation reversed.

In addition, it does not commit us to concluding that anyone actually takes the path. This must be deduced separately.

## 5 Computing Discourse Representation Structures

So far, we have only treated simple sentences without much of the surrounding context. The real test for this analysis is how it interacts with the constraints on interpretation posed by its surrounding context.

- (9) (...) nous descendons, pendant un quart d’heure, la vallée de l’Eséra.  
*we descend, for a quarter of an hour, the Es-era valley.*
- (10) La lune, qui éclaire notre marche, nous fait découvrir sur la droite un sentier qui serpente.  
*The moon, which lightens our steps, allows us to discover a winding path on our right.*
- (11) Il nous conduit sur un petit plateau, au milieu de sapins, au-dessus et à quelque distance du torrent de Ramun.  
*It leads us to a small plateau, surrounded by firs, at some distance of and above the Ramun torrent.*

Here, “Il” (it) in sentence (11) refers to “un sentier qui serpente” (a winding path), so correct analysis require resolution of the anaphor before coercion in order to give a correct analysis.

As is well known, the rhetorical relations as used by SDRT (Asher and Lascarides, 2003), provide a set of important constraints on the possible interpretations on discourse. In the exam-

ple above, we have the relations *Background*(9,10) and *Narration*(10,11).

A second example illustrates the importance of rhetorical structure on the interpretation (as well as the difficulty of automatically obtaining such a structure).

- (12) Nous partimes pour Barèges à 8 heures du matin par une fort jolie route qui nous conduisit à Lourdes.  
*We left (PS) for Barèges at 8 in the morning, taking a very pretty road which led (PS) us to Lourdes.*
- (13) (...) qui va en se resserrant jusqu'à Pierrefite, où les routes de Lux et de Cauterets séparent.  
*(...) which goes shrinking along the way, up to Pierrefite, where the roads to Lux and to Cauterets split.*
- (14) Celle de Lux entre dans une gorge qui vous mène au fond d'un précipice et traverse le gave de Pau.  
*The one to Lux enters a gorge which leads you to the bottom of a precipice and traverses the Gave de Pau.*
- (15) (...) Après une longue marche, l'on arrive à Barèges à 6 heures du soir.  
*(...) After a long walk, we arrive in Barèges at 6 in the evening.*

Here, sentence (12) introduces the destination and therefore the whole spatio-temporal extension route. The following will therefore constitute an Elaboration relation between this sentence and the sequence of (13)-(15). It is (at first sight) difficult to decide on the discourse relation of Sentence (14): it would certainly be possible to have a later phrase beginning with "Celle de Cauterets" (the road leading to Cauterets) and a number of the following sentences (omitted here for space reasons) give further background information about the road to Lux. However, at sentence (15), it suddenly becomes evident that the author has been describing the road while following it.

## 6 Conclusions and Future Work

We have given a treatment of "virtual movement" in a type-logical grammar. Our account merges two successful extensions of "standard" Montague-style semantics into a single, coherent type-theoretic framework.

## References

- Anne Abeillé, Lionel Clément, and François Toussenenel. 2003. Building a treebank for French. In Anne Abeillé, editor, *Treebanks: Building and Using Parsed Corpora*, chapter 10, pages 165–187. Kluwer, Dordrecht.
- Nicolas Asher and Alex Lascarides. 2003. *Logics of Conversation*. Cambridge University Press.
- Nicolas Asher, Philippe Muller, and Mauro Gaio. 2008. Spatial entities are temporal entities too: The case of motion verbs. In *LREC 2008 Workshop on Methodologies and Resources for Processing Spatial Language*, Marrakech, May 2008.
- Christian Bassac, Bruno Mery, and Christian Retoré. 2010. Towards a type-theoretical account of lexical semantics. *Journal of Logic, Language and Information*, 19(2):229–245.
- Johan Bos, Stephen Clark, Mark Steedman, James R. Curran, and Julia Hockenmaier. 2004. Wide-coverage semantic representation from a CCG parser. In *Proceedings of the 20th International Conference on Computational Linguistics (COLING-2004)*, pages 1240–1246, Geneva, Switzerland.
- Ray Jackendoff. 1983. *Semantics and Cognition*. MIT Press.
- Hans Kamp and Uwe Reyle. 1993. *From Discourse to Logic*. Kluwer Academic Publishers, Dordrecht.
- Pierre Loustau. 2008. *Interprétation automatique d'itinéraires dans des récits de voyages D'une information géographique du syntagme à une information géographique du discours*. Ph.D. thesis, Université de Pau.
- George A. Miller and Philip Johnson-Laird. 1976. *Language and Perception*. Harvard University Press.
- Richard Moot. 2010a. Semi-automated extraction of a wide-coverage type-logical grammar for French. In *Proceedings of Traitement Automatique des Langues Naturelles (TALN)*, Montreal.
- Richard Moot. 2010b. Wide-coverage French syntax and semantics using Grail. In *Proceedings of Traitement Automatique des Langues Naturelles (TALN)*, Montreal.
- James Pustejovsky. 1995. *The generative lexicon*. M.I.T. Press.
- Leonard Talmy. 1999. Fictive motion in language and "ception". In Paul Bloom, Mary A. Peterson, Lynn Nadel, and Merrill F. Garrett, editors, *Language and Space*, pages 211–276. MIT Press.